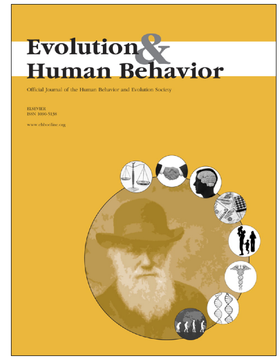
Accepted Manuscript



A Noisy Signal: To what extent are Hadza hunting reputations predictive of actual hunting skills?

Duncan N.E. Stibbard-Hawkes, Robert D. Attenborough, Frank W. Marlowe

PII:	S1090-5138(18)30005-9
DOI:	doi:10.1016/j.evolhumbehav.2018.06.005
Reference:	ENS 6216
To appear in:	*Evolution and Human Behavior*
Received date:	12 January 2018
Revised date:	6 June 2018
Accepted date:	10 June 2018

Please cite this article as: Duncan N.E. Stibbard-Hawkes, Robert D. Attenborough, Frank W. Marlowe , A Noisy Signal: To what extent are Hadza hunting reputations predictive of actual hunting skills?. Ens (2018), doi:10.1016/j.evolhumbehav.2018.06.005

A Noisy Signal: To what extent are Hadza hunting reputations predictive of actual hunting skills?

Duncan N. E. Stibbard-Hawkes[a,*], Robert D. Attenborough[b,c], Frank W. Marlowe[b]

[a]*Department of Anthropology, Durham University, Dawson Building, South Rd, Durham, DH1 3LE, UK*
[b]*Department of Archaeology and Anthropology, Cambridge University, Cambridge, CB2 1QH, UK*
[c]*School of Archaeology and Anthropology, A.D. Hope Building (#14), The Australian National University, Ellery Crescent, Acton, ACT 2601, Australia*

Abstract

The measurement of hunting ability has been central to several debates about the goals of men's hunting among the Hadza and other hunter-gatherer populations. Hunting ability has previously been measured indirectly, by weighing the amount of food individuals bring back to camp over an extended period, their central place hunting return rate, and by conducting hunting ability interviews. Despite the centrality of the hunting ability concept, some authors have expressed scepticism that such measures accurately capture individual differences in actual hunting ability. In the current study, we introduce a novel measure of hunting reputation which, unlike previous ones, allows fine-grained distinction between hunters of all reputations. To assess the suitability of this measure as a viable proxy for hunting ability, we address two further questions. First, to what extent do interviewees agree about the hunting ability of their present and former campmates? Second, to what extent does this measure of hunting reputation reflect success in four tasks expected to capture components of hunting ability? We demonstrate that measures of hunting reputation appear to capture actual variation in component hunting skills. We argue, however, that hunting reputation appears too noisy a measure of hunting ability for hunting to act, as some have suggested, as an honest signal of cryptic qualities related to hunting ability.

Keywords: Hunter-Gatherers, Hunting, Reputation, Foraging, Costly Signalling

Word Count (Calculated with TeXcount; Excluding Abstract and References and Acknowledgements):
Words in text: 8565 Words in headers: 82 Words outside text (captions, etc.):

*Corresponding author
Email address: duncanstibs@cantab.net (Duncan N. E. Stibbard-Hawkes)

1. Introduction

Although hunter-gatherer diets are highly variable (Kelly, 2013), there are no forager societies in the modern ethnographic record who do not practise hunting and/or fishing in some form (Murdock, 1967; Kelly, 2013). The earliest evidence of butchery (Domínguez-Rodrigo et al., 2005; McPherron et al., 2010) dates back to at least 2 million years ago [MYA], possibly as early as 3.9 MYA. The earliest evidence for weapon-assisted hunting dates back to perhaps as early as 800,000 YA (Rabinovich et al., 2008) and at least 400,000 YA (Thieme, 1997). For this reason, hunting and hunting ability have been central to many of the key debates in evolutionary anthropology over the past decade. Hunting and meat eating have been implicated in the evolution of the human capacity for cooperation (Bingham, 2000; Kaplan et al., 2007), human life history (Gurven et al., 2006; Kaplan et al., 2007), the sexual division of labour (Bird, 1999; Marlowe, 2007; Apicella et al., 2017) and the expansion of the human brain (Aiello & Wheeler, 1995). It has also been suggested by some (Hawkes & Bird, 2002; Smith, 2004; Smith & Bliege Bird, 2005) that hunting might act as an 'honest' or 'costly' signal of a hunter's skill, vitality, health or some other relevant but otherwise cryptic quality of interest to potential mates, rivals or allies.

Hunting ability among modern hunter-gatherer groups has been of great interest to evolutionary anthropologists, seeking to test hypotheses about hunting and its evolutionary origins (Hawkes et al., 1991; Hill et al., 1993; Kaplan & Hill, 1985; Alvard & Gillespie, 2004; Marlowe, 2005; Gurven & von Rueden, 2006; Apicella, 2014). These studies have approached research into hunting ability in a variety of ways, and it is important to consider the possible meanings of the concept in principle, and the ways in which it can be operationalised. 'True hunting ability' (i.e. true skill at finding and killing wild animals given equal luck and effort) probably cannot practicably be measured directly in most hunter-gatherer settings. Instead, the issue has been addressed by developing more readily implemented measures intended as proxies for true hunting ability. Their efficacy as proxies may be more or less precise and depends on certain assumptions.

To date there have been two main such approaches. Where there are sufficient data, central place hunting returns - the weight or caloric value of all hunted foods brought back to a central place per individual per day - are measured or estimated, and assumed to represent fairly well the hunter's success, and therefore (his, usually) hunting ability. Other studies, usually in the absence of such data, instead make the assumption that an individual's hunting ability is well and fairly accurately known to campmates and peers; and on that basis these studies use measures of hunting reputation as proxies for true hunting ability. Further, measures of hunting reputation are of interest in their own right: it perceived hunting ability (or reputation) itself, whether well or ill founded, that influences companions' behaviour and not actual ability or success. This is especially relevant where, as among the Hadza, hunting is usually undertaken alone. Both methods are also subject to the assumptions that success broadly reflects ability, i.e. time and effort put into hunting are similar across individuals (which is reasonable for Hadza hunters), and that sample size is sufficient to minimise the impact of varying luck.

Central place hunting return measures have been widely used in studies of hunter gatherer populations (e.g. Hawkes et al., 1991; Kaplan & Hill, 1985; Marlowe, 1999; Hawkes et al., 2001). Such methods are useful because they provide an exact numerical measure, in either kg or kcal, of each hunter's daily contribution, which can be easily employed in statistical testing. Furthermore, in at least some populations, central place hunting returns measures have been shown to correspond to estimates of hunting ability given by a hunter's peers and neighbours (e.g. Marlowe, 1999; Koster, 2010). The method does, however, have two major limitations.

Firstly, central place returns measures may not capture resources acquired and eaten away from a central place. This is especially important in those populations such as the Hadza where hunting is predominantly a solitary activity. A recent study of foods eaten by men on solo foraging trips (Berbesque et al., 2016) showed that, over 118 follows, men ate much food (mean = 2405 kcal/day) out of camp. Although honey accounted for a high percentage of these calories, much small game was also eaten away from camp during the early wet season. For this reason, central place return measures may, in part, capture proclivity to share, rather than skill at finding and killing animals.

A second problem is that many foragers live in environments where there is high daily variance in hunting returns (Winterhalder, 1986). Hill & Kintigh (2009) have demonstrated that, where variance in hunting success is high, a large sample of days is required to accurately estimate a hunter's true mean hunting return rate (Hill & Kintigh, 2009). Among Aché hunters, who bring home game on 48% of days, Hill & Kintigh (2009) calculated that a sample of 260 days is required to estimate a hunter's true mean daily acquisition rate with ±20% accuracy. Among the !Kung (Ju/'hoansi), who acquire meat on <27% of hunting days, a sample of >600 hunting days is required (Hill & Kintigh, 2009). Hadza hunters kill large game on between 3% Hawkes et al. (1991) and 0.97% (Wood & Marlowe, 2013; Blurton Jones, 2016) of days (approximately one a month and once in three months) while Berbesque et al. (2016) found that hunters killed any animal, large or small, on 16.44% of days (approximately five kills per month). For this reason, the sample size required to accurately estimate mean daily hunting return rates among the Hadza is untenably large.

Where long-term data on individual hunting returns are unavailable, 'hunting reputation' is often used as a proxy for hunting ability instead. This has classically been measured, among the Hadza, using the 'nominations method' (e.g. Marlowe, 1999, 2000; Blurton Jones & Marlowe, 2002; Marlowe, 2010; Blurton Jones, 2016): Interviewees are asked to nominate two or three 'exceptional' hunters. A man's 'hunting reputation' is then defined as the number of times he has been nominated as an exceptional hunter as a proportion of all nominations. This method has several advantages: The task is easy to understand with minimal risk of confusion. Questions may be answered quickly and it is possible to ask several different questions - e.g. 'Who is exceptionally hard working?', 'Who is very good at digging?' - in one sitting. Answers are based upon every person each interviewee knows and are, therefore, generalisable across the whole population.

Although straightforward and efficient, the nomination method has some disadvantages. The most serious is that the 'majority of men... receive few or no nominations' (Blurton Jones, 2016, p.426). Blurton Jones (2016) refers to these men as the '"poor hunter" majority', with 'poor hunter' rightly in inverted commas; it is impossible to know whether non-nominated hunters are 'poor' or not. The nomination method cannot capture distinctions in hunting ability between poor and average hunters or even those who are good but not exceptional. It is only possible to use nomination data to answer questions about the differences between exceptionally good hunters versus 'the rest'. The second problem with the nomination method is that it does not account for the fact that some individuals, by dint of age, personality or regularity of residential moves, may be more widely known than others and, thus, more likely to be nominated regardless of actual hunting ability.

Informant-ranking, whereby respondents rank a list of people, may also be used to measure hunting reputation (see Koster, 2010; Apicella, 2014). Only one previous Hadza study (Apicella, 2014) has employed ranking-based methods. Apicella (2014) asked interviewees to rank only those hunters living in the same camp, rather than ranking all hunters in the study. This method was successful and demonstrated a clear relationship between within-camp hunting reputation and strength. Furthermore, unlike the nominations method, Apicella's study allowed distinctions to be made between hunters at all levels of ability.

The disadvantage of within-camp ranking, however, is that it only allows comparisons of hunters within camps/settlements. Comparisons cannot easily be made across camps - a problem as the best hunter in a camp of poor hunters may be equivalent in skill to the worst hunter in a camp of good hunters. Furthermore, as, among the Hadza, there are seldom more than 15 hunters living in any one camp, sample sizes for within-camp rankings are small. There may only be a small number of interviewees in each camp and interviewees of dissenting opinion may influence results excessively. Lastly, individuals often live with close friends as well as parents, siblings and spouses (Blurton Jones, 2016), who may introduce biases.

There are two further criticisms levied at hunting reputation measures in general. First, it is possible that, as Kelly (2013, p.223) suggests, interviewees may interpret the question 'Who is the best hunter?' as 'Who is a good all-around man?' Reputations may reflect other positive qualities such as generosity or personableness, rather than skill at finding and killing wild animals (hunting ability).

Second, Hill & Kintigh (2009) have argued that, just as anthropologists cannot reliably distinguish true daily hunting return rate without a very large sample of hunting days, so too may a hunter's neighbours, peers and campmates be unable to accurately assess his true hunting success. Indeed, Hill & Kintigh (2009) have argued that, in previous studies (e.g. Marlowe, 1999), hunting reputation measures are correlated with other measures of hunting ability due to 'only a few exceptionally good or poor hunters' (p.375).

This question is further important to the costly signalling hypothesis of human hunting. It has been suggested that the adaptive function of men's hunting is to signal some quality, otherwise cryptic to observers (Hawkes & Bird, 2002). Proponents of the costly signalling hypothesis of human hunting seldom make explicit exactly what these qualities are. When they do suggest what such qualities might be most are related to a hunter's skill at finding and killing wild animals: e.g. 'skill and cognitive ability' & 'physical vigour' (Bliege Bird & Smith, 2005, p.237); 'strength, skill' (Smith et al., 2003, p.118); 'ethological knowledge, visual acuity [or] stamina' (Smith & Bliege Bird, 2005, p.142). Proponents of the costly signalling hypothesis (Hawkes & Bird, 2002) also argue that food redistribution is the mechanism by which such qualities are signalled so that campmates make a generalised assessments of hunting ability and any relevant associated cryptic qualities based on the food they receive. If the adaptive function of men's hunting is indeed to signal qualities associated with hunting ability to campmates, then, it is important that campmates A) are in general agreement about hunters' abilities and B) that measures of hunting reputation correspond to the actual skills that make up hunting ability. If either of these conditions were not met, for example because variance in food acquisitions (i.e. 'luck') obfuscates a hunter's actual abilities (Hill & Kintigh, 2009), or because unrelated qualities such as personableness have undue influence on a hunter's perceived ability (Kelly, 2013), doubt would be cast on the efficacy of hunting as a reliable honest signal.

In the current study we attempt to address each of these concerns. We introduce a new measure of hunting ability based on interview ranking of an entire sample. This allows granular distinctions to be made between hunters at all levels of reputation, rather than just those perceived to be the best. Furthermore, unlike Apicella's (2014) method, results are generalisable across camps. Data are high resolution enough to enable us to assess the extent of agreement between independent interviewees both concerning the sample as a whole and concerning those hunters of 'middling' reputation. We also relate hunting reputations to success at tasks designed to test four component hunting skills.

We employed four measures: Aim with a bow and arrow; upper-body strength; ability to recognise animal vocalisations; and eyesight. These were chosen to represent a substantial subset of those component hunting skills which contribute to hunting ability. Unlike other proposed component hunting skills such as stealth, spotting and tracking (Blurton Jones, 2016, p.281), these can be practically and expediently measured. However, each measure captures an important component of hunting ability: aim is important because animals flee when ambushed and a hunter may have only one opportunity to shoot per encounter. Upper-body strength has been elsewhere shown to be significantly positively correlated with a hunting reputation measure (Apicella, 2014) and is of clear utility in many foraging tasks such as climbing and shooting with bows. Most animal species produce calls or other vocalisations that are unique to that species. Knowing these calls is important when locating prey because much of the Lake Eyasi region is bushland with little open grassland (Blurton Jones, 2016, p.13) and vision is often obstructed. Finally, visual acuity is important because many hunting-related tasks - spotting movement, aiming, identifying tracks and other spoor - are vision dependent.

2. Materials and Methods

2.1. Study Population

We conducted research among the Hadza, an ethno-linguistic group living in the Eyasi Region of Northern Tanzania. There are ∼1000 Hadza speakers in the region (Blurton Jones, 2016). Due to increasing use of domesticated grains, the oft-cited statistic that '250 [Hadza] continue to hunt and gather with traditional technologies for approximately 95% of their total diet' (Wood & Marlowe, 2014, p.540) appears, based on personal observation, an overestimate. It is probable, however, that this many Hadza still rely on foraging for a large majority of their diet. The Hadza are one of only a small number of hunter-gatherer populations in East Africa and have been foraging in the region for as long as there have been written records (Obst, 1912; Bagshawe, 1925). In numerous ways the Hadza are close to the median of all warm-climate forager groups (Marlowe, 2010, p.261).

Meat from hunted game, small and large, is estimated to provide ∼32% (Marlowe et al., 2014) of all calories in the Hadza diet. The Hadza also have a strong sexual division of labour and almost all hunting is carried out by men. Most meat, especially that from large game species, is brought back to a central place and widely shared (Marlowe, 2010, p.103), although some small game is eaten outside of camp (Berbesque et al., 2016). Most Hadza hunting trips (89%) are solitary (Berbesque et al., 2016), although hunters will occasionally cooperate in night-time water-hole ambushes during the dry season, when ground water is scarce and certain prey species are more clustered (Marlowe, 2010, p.118). Men hunt with bows and poisoned arrows. In the majority of cases large game are shot and followed until they succumb to poison-induced cardiac arrest. The Hadza hunt over 880 species of animal, the majority (>750) of which are birds (Marlowe, 2010, p.127). Hadza hunters have been described as 'big game specialists' (Hawkes et al., 2014, p.607), although the extent of their reliance on large game has recently been the subject of some debate (Wood & Marlowe, 2013; Hawkes et al., 2014; Wood & Marlowe, 2014).

Studies of the Hadza have been at the forefront of discussions about the adaptive value of human hunting (e.g. Hawkes et al., 1991; Marlowe, 1999; Wood & Marlowe, 2013; Hawkes et al., 2014; Wood & Marlowe, 2014; Blurton Jones, 2016), including but not limited to the costly signalling hypothesis. Therefore, although only one of many modern forager groups, the Hadza are an ideal study population to address the efficacy of hunting reputation measures.

2.2. Field Trips and Sample Size

This study was conducted over three field trips to Tanzania, each roughly a month long. The dates of these were: 17th August -17th September 2013 (mid dry season); 7th December 2013 - 6th of January 2014 (early wet season);

19th October 2014 - 27th of November 2014 (late dry season, beginning of wet season). Over the course of the project we visited a total of 17 independent Hadza camps. The 72 hunters (all male) who made up the final dataset were drawn from nine camps and 89 interviewees (male and female) were drawn from 13 camps with some overlap.

2.3. Demographic and Anthropometric Data

During each camp visit we collected basic demographic and anthropometric data from all participants. Data included age, weight, height, place of birth, name of both parents and, where appropriate, name of current spouse, number of children born, number of children still living and names of children living in camp.

2.3.1. Age Estimation

Those Hadza who have attended school usually know their age, as do many under the age of 45 years. Age could not be verified by any official documents, although participants were not afraid to admit when they did not know their age. Many older Hadza do not know their exact ages and some could not give a decade, a recurring problem in studies of non-literate populations (Howell, 1979; Hill & Hurtado, 1996). Frank Marlowe has been working with the Hadza for over 20 years and his PhD supervisor Nicholas Blurton Jones for over 35. They have collected demographic data and age estimates from field trips throughout this period. Where available, we compared ages in our dataset to those given by Marlowe/Blurton Jones for the same hunters. Where hunters did not know their own age, or where there was significant disagreement, we used these ages. In those few cases where a hunter was not listed in previous demographic data and had not provided an age, we estimated the age of the hunter visually.

2.4. Hunting Ability Measures

For each of the 72 hunters in our study, we collected data on four metrics we believe likely to represent independent or semi-independent components of hunting ability. These were: aim with a bow and arrow; strength; ability to recognise animal vocalisations; and visual acuity. In each case there were ∼four dropouts or people who did not participate.

In line with the methods of Blurton Jones & Marlowe (2002), aim with a bow and arrow was measured in an archery contest. We created a 61cm by 61cm cardboard target with an outer circle of 33cm in diameter, an inner circle of 20cm in diameter and a 'bullseye' of 4cm in diameter. We ensured that the terrain was flat and the target was unobscured. It was impossible to control for moment to moment fluctuations in wind direction. Furthermore, we did could not estimate exact wind speed. Instead we used the Beaufort scale to estimate wind speed and always conducted the test at a Beaufort wind force of two or less, corresponding to wind speeds of $\leq$7mph. Hunters took three shots with their own bows from 10, 20 and 30m distance, nine shots in total. We awarded 25 points if the archer hit the outer ring, 50 points for the middle ring and 100 points for a bullseye. Sixty four of 72 hunters participated in this measure.

We measured bow pull strength using an Easton Digital bow pull scale. Hunters were asked to hook the scale to the string of their bow and pull back with as much strength as they could for fifteen seconds using their dominant arm. We recorded weight at peak pull in lbs. Sixty eight of 72 hunters participated in this measure

To measure knowledge of animal vocalisations we located recordings of 21 animals (Table 1) known to be hunted by the Hadza (Marlowe, 2010, p.127) or animals known to be relevant to hunting or foraging such as hyaena and leopards, from whom the Hadza may scavenge kills, or honey guides, which lead men to bees' nests. We took audio clips from various online animal sound repositories including the *Cornell University Macauley Library*, the *British Library Sound Archive*. We collected several audio clips of each species and cross-referenced each clip to ensure we had chosen a common vocalisation. The clearest recording of each was edited into a short clip of between 5 and 10 seconds each and clip volume was normalised. Each hunter was played each of the 21 clips three times and asked to identify the animal in either Swahili, English or the Hadza language. Correct guesses were scored 1 and incorrect guesses were scored 0. 'Close' guesses (for example, cow instead of buffalo, as indicated in Table 1) were scored 0.5. Seventy one of 72 hunters participated in this measure.

Hunters' visual acuity was tested using a 3m Landolt C Optotype which does not require knowledge of the Latin alphabet (see Keeffe et al., 1996). Under field conditions, it was impossible to ensure uniform lighting conditions (Durst et al., 2011). However, before commencing each test, we ensured that sun was behind the participant and that that the chart was clearly illuminated and free from glare. Results were recorded as LogMAR scores. Seventy one of 72 participants participated in this measure.

2.5. Hunting Reputation Interviews

After we had collected a dataset including 72 hunters, we conducted hunting ability interviews. We interviewed every willing person over the age of 16 in 13 different camps in 3 different regions - a total of 89 interviewees of both sexes. Interviews were conducted in Swahili. Interviewees were invited to a location where they could not be overheard and were asked to identify 72 high-quality A6 (105mm x 148mm) laminated colour prints of the faces of every hunter in our dataset. For each photograph, we asked the following questions:

1. What is the name of this person?
2. What is the name of his father?
3. When did you last live in the same camp as this person?

Where the interviewee did not recognise or had not lived with a particular hunter within the last two years, or did not know that hunter's name or the name of the hunter's father, we removed that photograph. Some Hadza use several names and some had absent fathers or multiple step fathers. Where the name or father's name provided was unfamiliar to us, we did not remove

Table 1: Species used in animal call recognition test with common name, Latin name and Hadza name provided. We here provide the Latin names reported by Marlowe (2010), although other Latin names are in use in some cases.

Species Name	Latin Name	Hadza Name
Spotted Hyaena*	*Crocuta crocuta*	Udzameko
Greater Honeyguide*	*Indicator indicator*	Tik'iliko
Helmeted Guineafowl*	*Numida meleagris*	Ch'aako
Lion	*Panthera leo*	Seseme
Olive baboon	*Papio anubis*	Ne'e'ko
Buffalo†	*Syncerus caffer*	Nakomako
Senegal Bushbaby**	*Galago senegalensis*	Chacha
Brown Greater Galago**	*Otolemur crassicaudatus*	Ndonoko
Tree Hyrax***	*Dendrohyrax arboreus*	Chasho
Rock Hyrax***	*Procavia johnstoni*	Ch'abako
Side-Striped Jackal‡	*Canis adustus*	Mołola
Vervet Monkey	*Cercopithecus aethiops*	Numbili
Impala	*Aepyceros melampus*	Popoako
Black Stork	*Ciconia nigra*	Gwengweyako
Marabou Stork	*Leptoptilos crummeniferus*	Nyambulu
Grey Crowned Crane	*Balearica regulorum*	Owania
Saddle-bill Stork	*Ephippiorhynchus senegalensis*	!Tatamu
Sacred Ibis	*Threskiornis aethiopicus*	Gijiko
Black-Crowned Night Heron⋆	*Nycticorax nycticorax*	Bungitape
Greater Flamingo	*Phoenicopterus roseus*	Gogo gogo
Collared Pratincole	*Glareola pratincola*	Qela qetape

* The vocalisations of these species are well known and were included to ensure that participants could properly interpret the recordings.
** The Senegal bushbaby was often mistaken for the brown greater galago and vice versa.
*** The rock hyrax was often mistaken for the tree hyrax and vice versa.
† Buffalo vocalisations were often mistaken as those of domesticated cattle, *Bos taurus*.
‡ One person mistook the vocalisations of the side-stripe jackal for those of a domesticated dog *Canis lupus familiaris*.
⋆ No hunters correctly recognised the maribou stork, although 15 hunters said that the recording was of the black-crowned night heron. As the vocalisations of the two species were phonically similar (despite striking differences in appearance), we counted these as 'close' guesses.

the photograph but sought clarification in later interviews. If we could not find later corroboration we removed that hunter from the ranking *post hoc*. This may have lead to some hunters being incorrectly labelled as unknown by particular interviewees, though we felt that less inclusive data were probably sounder. One hunter was only known by eight other people and was, for this reason, removed from the final sample, leaving 71 hunters in total.

We set out the remaining photographs in a random order. Interviewees were then asked to remove the photograph of the best hunter. By removing photographs one-by-one, interviewees ranked all the hunters they knew and had lived with in the last two years in order of hunting ability. We removed lists that were too short (<20 individuals) or were otherwise deemed unreliable, leaving a total of 67 ranked lists in our final sample.

2.6. Aggregated Hunting Reputation Measure

The final dataset included 71 hunters (labelled H_1, H_2, H_3, ..., H_{71}) and 67 ranked lists. For each hunter we calculated an aggregated hunting reputation score by taking the mean of the hunter's position in every list in which he appeared. The method used to calculate a hunter's hunting reputation was as follows: Let L_j denote the length of the jth ranked list (i.e. the number of names on the list). As no interviewee knew all hunters, in this study $20 \leq L_j \leq 69$. We define the *relative rank* $R_{i,j}$ of hunter H_i on list j by setting

$$R_{i,j} = \frac{L_j - \text{the position of } H_i \text{ in the order of list } j}{L_j}$$

This means that if H_i is at the top of list j, his relative rank $(L_i - 1)/L_i$ is close to 1 (between 0.95 and 0.99 in this study depending on the size of L_j) and is marginally higher in longer lists, whereas at the bottom his relative rank is zero. A hunter's relative rank — his 'score'— is the proportion (or fraction) of the way up a list his name appears. For example a hunter near the middle of a list of length 50 scores ∼0.5, which compares sensibly with the zero score of a hunter at the bottom of a list of length 20.

As no hunter appears in every ranked list, we take their average relative ranks over just those lists they do appear in. Setting $R_{i,j} = 0$ when hunter H_i does not appear in list j allows us to define this average — the *mean hunting score* (i.e. the mean relative rank), μ_i of hunter H_i — as follows:

$$\mu_i = \frac{R_{i,1} + R_{i,2} + R_{i,3} + \cdots + R_{i,67}}{\text{the number of lists in which } H_i \text{ appears}} . \qquad (\alpha)$$

In what follows we refer informally to the values of these mean hunting scores $\mu_1, \mu_2, \ldots, \mu_{71}$ as 'hunting reputation'.

2.7. Renown Measure

While only 67 interviewees provided ranked lists of hunting ability meeting the criteria for inclusion in an aggregated hunting rank (see Section 2.6), 89 interviewees answered questions about whether or not they knew the listed hunters. An interviewee was deemed to 'know' a hunter if they could correctly recall that hunter's first name and their father's name. For each hunter, we summed the number of interviewees who were deemed to know them. We define this number as a hunter's 'renown'.

3. Results

3.1. Agreement Between Interviewees

3.1.1. Pairwise Comparison of Randomly Generated Subsets

To quantify the extent of agreement among interviewees, we split the set R of 67 ranked lists randomly into two disjoint subsets S and $\bar{S}$ containing 34 and 33 ranked lists respectively. Once the 34 lists in S were chosen, $\bar{S}$ contained the 33 remaining lists not in S. Using Equation (α), where we applied it to the set R, we calculated the *mean hunting score* of the hunters independently for each of the subsets S and $\bar{S}$ in turn. These sets of ranks are denoted by

$$\mu(S) = \{\mu(S)_i \,:\, 1 \leq i \leq 71\} \text{ and } \mu(\bar{S}) = \{\mu(\bar{S})_i \,:\, 1 \leq i \leq 71\}$$

Spearman's rank correlation is an appropriate test for the level of agreement between two ordered lists of the 71 hunters ranked according to their $\mu(S)$ and $\mu(\bar{S})$-values. However, the number of distinct partitions of R into disjoint subsets S and $\bar{S}$ (with 34 elements in S) is equal to

$$N = \frac{67!}{34!33!} \approx 1{\cdot}4 \times 10^{19}, s$$

and so it is not feasible to calculate a Spearman's correlation between $\mu(S)$ and $\mu(\bar{S})$ for *all* possible partitions. Instead we use a Monte Carlo-like method of random sampling. We wrote a simple script using the R programming language that randomly generates a given number, z, of pairs of disjoint subsets, S and $\bar{S}$, and we calculated the Spearman rank correlation coefficient for the corresponding ordered lists determined by $\mu(S)$ and $\mu(\bar{S})$. We set the value of z to 10,000, a value small enough to be easily computed but large enough to consistently give near-identical outputs when iterated. We can not guarantee that all the partitions were distinct although the chance of generating two the same was vanishingly small (less than one in 700 trillion).

For each choice of the subset S of R, we obtained two ranked lists of the hunters from the values of $\mu(S)$ and $\mu(\bar{S})$; we denote their Pearson rank correlation coefficient by ρ_S. In 10,000 trials, the mean value of ρ_S at 0.89 was high, ranging from $\rho_S = 0.78$ to $\rho_S = 0.95$ with standard deviation equal to 0.02. Even the most dissimilar of the 10,000 pairs demonstrated a high level of agreement. Running the same script using the Pearson product-moment correlation method gave similar results (mean $r = 0.89$, SD $= 0.02$, range $= 0.80$ - 0.92). Not every one of the 10,000 iterations, however, met the assumptions of the test, namely that the data follow a normal distribution.

3.1.2. Exhaustive Pair-Wise Spearman's Correlation Analysis

The method presented above provides a good indication of agreement in the whole sample but does not provide any information about sub-populations among rankers and does not make clear whether or not there are any clusters of rankers who disagree with the majority. To address this question we performed an exhaustive pair-wise Spearman's correlation analysis.

For each of the 67 ranked lists in our sample, we performed a Spearman's rank order correlation with each of the 66 other ranked lists, excluding from each pair those hunters who did not appear in both lists. A representation of results for all 4422 (2211 unique) correlation tests is provided in Figure 1.

The majority of correlations (87%) are positive indicating general consensus. However, Figure 1 reveals a 'contrarian' subset of eight (12% of) interviewees whose results are, on average, negatively correlated with other interviewees. 'Contrarians' generally disagreed with each other as well as with the majority. The contrarian group does not come from any particular geographic area nor age group. All but one of them are, however, female. The cause of this pattern is unclear. We noticed that female interviewees were more likely to grow bored with the ranking procedure and it is possible that the eight contrarians were ranking hunters at random. It is, however, also possible that the eight contrarians took the test seriously and that their results represent genuine disagreement. From the available data, we have no way of distinguishing between these two possibilities. For this reason, we did not exclude the contrarian group from later analyses.

3.2. Agreement over Exceptional and Average Hunters

Hill & Kintigh (2009) have argued that 'most reported associations between informant rank and measured hunting returns are statistically significant only because of a few exceptionally good or poor hunters' who 'stand out easily' (p.375) and that little is known about the hunting ability of the majority.

If Hill & Kintigh's argument applies, there should be greater consensus between interviewees about the ability of those hunters with the best and worst hunting reputations than about those hunters of middling reputation. We define the best and worst hunters, in Hill & Kintigh's words *'exceptional'*[1], as those 27 hunters with $\mu_i > 1$ SD from the mean of $M(A)$ setting $M(A)$ as the complete set of all *mean hunting scores* for hunters $H_{1,2,\ldots,71}$ derived from the complete sample of 67 ranked lists (Section 2.6). We define *middling* hunters as those 44 hunters with $\mu_i \leq 1$ SD from the mean of $M(A)$.

As in Section 3.1.1, we split our 67 ranked lists into z pairs of randomly generated disjoint subsets, S and $\bar{S}$, of 34 and 33 ranked lists each. As before, we generated $\mu(S)$ and $\mu(\bar{S})$ for z = 10,000. We divided each instance of $\mu(S)$ and $\mu(\bar{S})$ into those hunters who were > 1 SD and ≤ 1 SD from the mean of $M(A)$. We label the scores of the *exceptional* group $\mu(SE)$ and $\mu(\bar{S}E)$ and

[1] Here not used to denote exceptionally good hunters, but those hunters, relatively good and bad, in the upper and lower tails of the distribution.

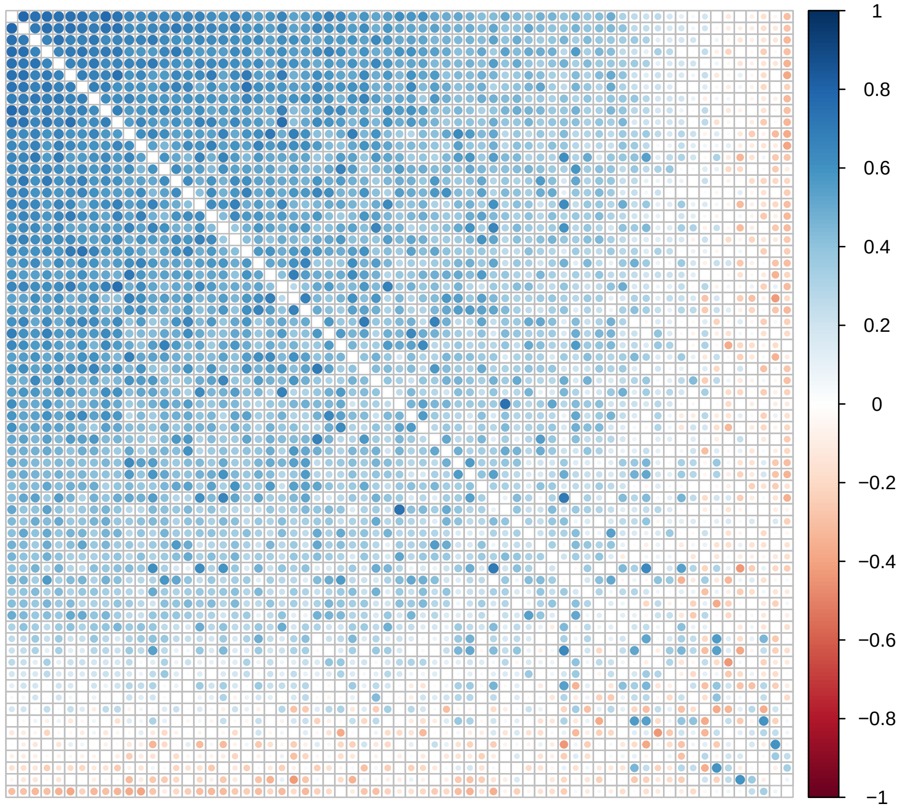


Figure 1: Matrix of correlations between all 67 ranked lists, mirrored on the horizontal axis and ordered by first principal component. Colour indicates the strength of the relationship, as indicated in the legend. The first row from the top and the first column from the left correspond to the same ranked list, as do the second from the top and second from the left etc.

the scores of the *middling* group $\mu(SD)$(and $\mu(\bar{S}D)$. For each instance of the 10,000 disjoint subsets we compared the *exceptional* hunters $\mu(SE)$ and $\mu(\bar{S}E)$ and then compared the *middling* hunters $\mu(SD)$ and $\mu(\bar{S}D)$ using the Spearman's correlation test.

Over 10,000 iterations, the mean ρ value for pairs from the exceptional group was 0.93 (SD = 0.02, range = 0.76-0.99). The mean ρ value for pairs from the middling group was lower at 0.85 (SD = 0.04, range = 0.69-0.96). This result confirms that there is more agreement between interviewees about hunters in the exceptional group than there is about middling hunters. However, agreement about both groups is high.

3.3. Hunting Reputation, Age and Hunting Ability

The relationship between age and mean hunting score in this study is noisy, with great variance in mean hunting scores at all ages. There is no significant linear relationship between age and hunting reputation (Table 2). Including age^2 improves the fit of the model and hunting reputation shows a weak but significant quadratic relationship with age (Table 2; Figure 2). Mean hunting scores in the present study peak between ages 40-55. For this reason, in later tests results, age and age^2 were included as a control.

Mean score in the archery task at 10m was 98. All but three participants managed to hit the target at least once from this distance. Accuracy declined considerably at greater distances. At 20m, participants achieved a mean score of only 34.5, ∼35% of the 10m scores. The number of people who failed to hit the target at all also increased from 3/67 at 10m to 22/66 at 20m. At 30m, only 28% of archers hit the target and mean score across all archers was 11.7 points, 12% of the mean score at 10m. Whether or not age and age^2 were included in the model, aim at all distances was a significant linear predictor of mean hunting score (Table 3). With each reduction in distance from the target, the predictive power of aim increased and the strongest predictor of a hunter's mean hunting score was aim at 10m only.

Mean draw strength was 61.61 lbs (28 kg), with a range of between 18.9 and 92.8 lbs (8.6 - 42.1 kg) and a standard deviation of 16.18 (7.3 kg). Draw strength shows a clear positive relationship with mean hunting score (Figure 2). This relationship is highly significant whether or not age and age^2 are included (Table 4).

Each hunter scored between four and 14.5 (mean = 10.54, amounting to ∼50% of the 21 recorded animal vocalisations) on the animal vocalisation recognition test. The success rate for recognising calls varied from almost 100% for species such as guinea fowl (68/70 hunters), and hyaena (67/70), to zero or near zero in the case of some of the less commonly sighted birds (e.g. sacred ibis, 0/70; saddle back stork 1/70). There is a clear positive relationship between hunters' mean hunting scores and the number of animal vocalisations that they were able to recognise (Figure 2). This relationship was significant both with and without age and age^2 included in the model (Table 5)

The patterning of Hadza distance vision was unremarkable. Mean visual acuity for the better eye was -0.03 (n = 70) and 0.04 (n = 65) for both eyes,

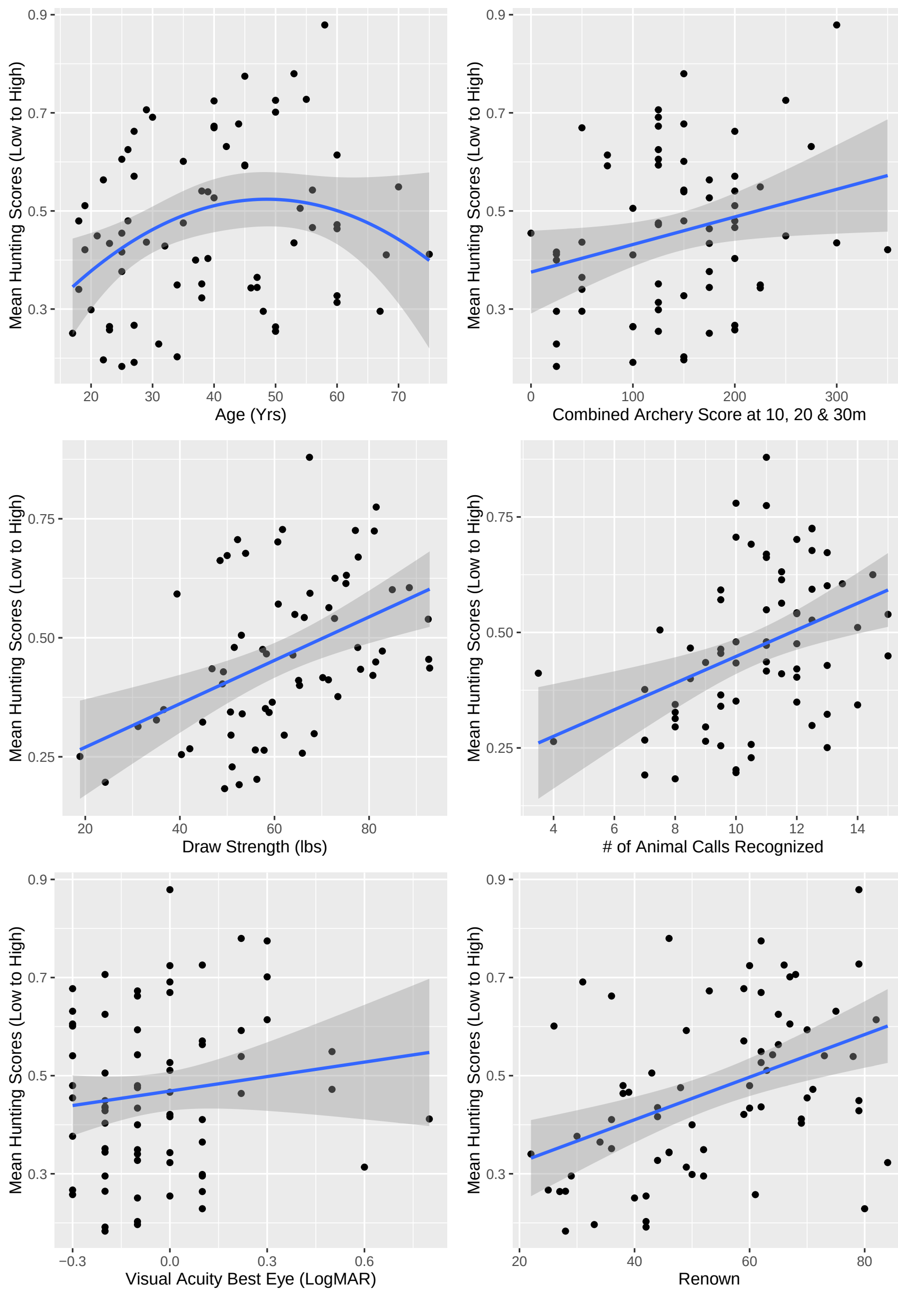


Figure 2: Scatterplots of Mean Hunting Score by A) Age, B) Archery Score, C) Draw Strength, D)Animal Vocalisation Recognition Score and E) Visual Acuity, overlaid with regression lines and 95% confidence bands.

equivalent to just under and just over 20/20 vision respectively. Better eye visual acuity ranged from -0.3 to 0.8, equivalent to 20/10 and 20/126 respectively. Visual acuity showed the expected linear decline with age (Visual Acuity (LogMAR) = 0.01(Age); p <.001; $R^2 = 0.41$). Better eye visual acuity did not show the expected relationship with *mean hunting score* (Figure 2) - better hunters had marginally worse vision. The relationship between mean hunting score and visual acuity was non-significant with and without age and age^2 included as a control (Table 6).

3.3.1. Model Selection

When aim at 10m, draw strength and animal vocalisation recognition were included in the same model (Table 7), controlling for age and age^2, each made a significant independent contribution to *mean hunting score*. Variance in age, draw strength, aim at 10m and knowledge of animal calls alone accounted for 44% of the total variance in hunting reputation. This model is only one of many models that may be derived from these parameters. Each additional variable artificially increases fit in a multiple linear regression and without running a large number of analyses it is difficult to tell whether any one variable is superfluous. To address this, we ran an all-subsets model selection (an exhaustive alternative to 'traditional' stepwise model selection) to choose the best model. We used the Akaike Information Criterion [AIC] to select, of the many possible models derived from our data, that which best explained variation in hunting reputation using the fewest number of extraneous variables. As each variable had a small number (1-6) of dropouts and missing cases, to conduct this analysis, nine hunters were removed and the total sample size was reduced to 62. The eight best candidate models are listed in Table 8.

The best fitting model included age and the same three variables identified as significant predictors in Section 3.3. The second best fitting model included age and age^2 and was similarly good ($\Delta_i \leq 1.5$). Those models that included better eye vision were substantially worse ($\Delta_i \geq 2$) than the best fitting model.

3.4. Excluding Highest- and Lowest- Ranked Hunters

To test whether interviewees were able to assess accurately only the best and the worst hunters, we excluded those hunters who scored above ($n = 5$) or below ($n = 4$) 1.5 standard deviations from the mean. When the four 'best' (see Section 3.3.1) predictor variables were included in a multiple linear regression, only two remained significant (Table 9). We next removed those hunters who scored above ($n = 13$) and below ($n = 14$) one standard deviation of the mean (those defined as 'exceptional hunters' in Section 3.2). Though animal vocalisation recognition approached significance, only bow pull strength remained significant at the 0.05 level. The adjusted value of R^2 was nearly halved when compared to the full sample, as was the amount of variation in hunting reputation explained by aim (see Table 7).

Table 2: Linear and Quadratic Regression Models of Mean Hunting Score and Age (yrs).

Step		B	SE B	β	R^2	Sig.
1					0.04	0.09
	Constant	0.38	0.06			0.00
	Age (yrs)$^{\cdot}$	0.002	0.001	0.21		0.09
2					0.10	0.03
	Constant	0.10	0.02			0.49
	Age (yrs)*	0.02	0.007	1.54		0.02
	Age^2*	-0.0002	0.00009	-1.35		0.04

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot$ < 0.1

Table 3: Linear and Quadratic Regression Models of Mean Hunting Score and Aim, controlling for the effects of Age and Age^2

Distances		B	SE B	β	R^2	Adj. R^2	Sig.
Aim at 10, 20 & 30m					0.14	0.10	0.02
	Constant	0.08	0.15				0.59
	Aim at 10-30m*	0.0006	0.0003	0.27			0.03
	Age (yrs)$^{\cdot}$	0.01	0.007	1.32			0.06
	Age^2 $^{\cdot}$	-0.0001	0.00008	-1.16			0.10
Aim at 10 & 20m					0.16	0.11	0.02
	Constant	0.09	0.14				0.54
	Aim at 10-20m*	0.0007	0.0003	0.29			0.02
	Age (yrs)$^{\cdot}$	0.01	0.007	1.21			0.08
	Age^2	-0.0001	0.00008	-1.04			0.13
Aim at 10m only					0.22	0.18	0.00
	Constant	0.10	0.14				0.48
	Aim at 10m**	0.001	0.0004	0.38			0.00
	Age (yrs)	0.01	0.007	1.04			0.12
	Age^2	-0.0001	0.00008	-0.86			0.20

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot$ < 0.1

Table 4: Linear and Quadratic Models of Mean Hunting Score and Draw Strength (lbs), including and excluding Age (yrs) and Age^2

Step		B	SE B	β	R^2	Adj. R^2	Sig.
1					0.19	0.18	0.00
	Constant	0.19	0.07				0.02
	Draw Strength (lbs)***	0.01	0.001	0.44			0.00
2					0.30	0.27	0.00
	Constant	-0.03	0.15				0.14
	Draw Strength (lbs) ***	0.005	0.001	0.45			0.00
	Age (yrs)**	0.02	0.007	1.61			0.01
	Age^2*	-0.0002	0.00008	-1.41			0.02

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot$ < 0.1

Table 5: Regression Model of Mean Hunting Score and Animal Vocalisation Recognition Score (/21), including and excluding for Age (yrs) and Age^2

Step		B	SE B	β	R^2	Adj. R^2	Sig.
1					0.16	0.15	0.00
	Constant$^{\cdot}$	0.16	0.09				0.07
	Animal Recognition (/21)***	0.03	0.01	0.40			0.00
2					0.27	0.24	0.00
	Constant	-0.19	0.15				0.22
	Animal Recognition (/21)***	0.03	0.008	0.45			0.00
	Age (yrs)$^{\cdot}$	0.01	0.007	1.10			0.07
	Age^2	-0.0001	0.00008	-0.80			0.19

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot$ < 0.1

Table 6: Regression Model of Mean Hunting Score and Better Eye Visual Acuity (LogMAR), including and excluding for Age (yrs) and Age^2

Step		B	SE B	β	R^2	Adj. R^2	Sig.
1					0.02	0.003	0.26
	Constant	0.47	0.02				0.00
	Better Eye Visual Acuity (LogMAR)	0.10	0.09				0.26
2					0.11	0.07	0.05
	Constant	0.10	0.14				0.54
	Better Eye Visual Acuity (LogMAR)	0.14	0.11				0.21
	Age (yrs) *	0.02	0.008				0.01
	Age^2*	-0.0002	0.0001				0.02

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot < 0.1$

Table 7: Two best Models of Mean Hunting Score, missing cases removed

Step		B	SE B	β	R^2	Adj. R^2	Sig.
1					0.44	0.40	0.00
	Constant $\cdot$	-0.17	0.10				0.09
	Aim at 10m**	0.001	0.0003	0.28			0.01
	Animal Recognition (/21)*	0.02	0.01	0.33			0.01
	Draw Strength (lbs)**	0.003	0.001	0.34			0.00
	Age (yrs)**	0.003	0.001	0.28			0.01
2					0.45	0.40	0.00
	Constant	-0.28	0.14				0.05
	Aim at 10m*	0.0008	0.0003	0.27			0.01
	Animal Recognition (/21) *	0.02	0.008	0.31			0.02
	Draw Strength (lbs)**	0.003	0.001	0.35			0.00
	Age (yrs)	0.01	0.006	0.94			0.11
	Age^2	0.00008	0.00007	-0.67			0.25

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot < 0.1$

Table 8: 13 Best Models of Hunting Reputation chosen using the Akaike Information Criterion

	Model	AIC Δ_i	w_i	acc w_i	Adj R^2
1	Age + Aim + Recog + Str	0.00	0.29	0.29	0.40
2	Age + Age^2 + Aim + Recog + Str	1.15	0.16	0.45	0.40
3	Age^2 + Aim + Recog + Str	1.34	0.15	0.59	0.39
4	Age + Aim + Recog + Vision + Str	2.54	0.08	0.67	0.39
5	Age + Age^2 + Aim + Recog + Vision + Str	3.22	0.06	0.73	0.40
6	Age^2 + Aim + Recog + Vision + Str	3.88	0.04	0.77	0.38
7	Age + Recog + Str	4.64	0.03	0.80	0.34
8	Aim + Recog + Vision + Str	4.86	0.03	0.83	0.35

Age = Age (yrs)
Aim = Archery Contest Score at 10m
Str = Bow Pull Strength (lbs)
Recog = Animal Vocalisation Regognition Test Score (/21)
Recog = Animal Vocalisation Regognition Test Score (/21)
Vision = Better Eye Visual Acuity (LogMAR)

Table 9: Best model with exceptional hunters removed.

Hunters Removed	Parameters	B	SE B	β	R^2	Adj. R^2	N	Sig.
Reputation $> \sigma 1.5$					0.30	0.25	55	0.00
	Bow Pull Strength (lbs)*	0.003	0.001	0.32				0.01
	Aim at 10m	0.0005	0.0004	0.20				0.13
	Animal Vocalisation Recognition (/21)*	0.002	0.008	0.33				0.02
	Age (yrs)	0.002	0.001	0.21				0.11
Reputation $> \sigma$					0.32	0.24	40	0.01
	Bow Pull Strength (lbs)*	0.002	0.001	0.41				0.02
	Aim at 10m	0.0003	0.0003	0.16				0.32
	Animal Vocalisation Recognition (/21)$^{\cdot}$	0.001	0.007	0.29				0.10
	Age (yrs)	0.001	0.0009	0.20				0.21

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot < 0.1$

3.5. Hunting Reputation and Renown

There is no significant relationship, either linear or quadratic, between renown and age (Table 10); and some of the best known hunters were <40 years old. Renown showed a significant but weak positive linear relationship to hunting reputation (Table 10, Figure 2).

Table 10: Regression Models of Factors Influencing Renown

		B	SE B	β	R^2	Adj. R^2	Sig.
1	Age (yrs)	0.10	0.14	0.09	0.01	0.01	0.46
2	Age (yrs)	0.70	0.77	0.62	0.02	-0.01	0.37
	Age^2	-0.01	0.01	-0.54			0.43
3	Mean Hunting Score***	43.09	10.81	0.43	0.19	0.18	0.00

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$
$\cdot < 0.1$

4. Discussion

4.1. Hunting Reputation: Agreement and Validity

Interview-based hunting reputation measures are only useful if they fulfil two criteria. First, interviewees ought generally to agree. Second, if mean hunting score reflects true hunting ability, then it should be positively correlated with success at tasks which measure important component hunting skills.

We have demonstrated, using two different measures, that independent interviewees show a high level of agreement. We identified a small 'contrarian' subgroup of eight individuals who, on average, disagreed with the majority. Furthermore, agreement is somewhat higher over exceptional hunters and somewhat lower over average hunters. In all cases, however, overall agreement was high (Section 3.2). This is important as it suggests that interviewees A) understand the task and B) are rating hunters based on a similar set of criteria. In previous studies it has generally only been assumed, rather than demonstrated, that this was the case, and these results go some way towards justifying the past and future use of reputation-based measures as a proxy for true hunting ability.

Such general agreement is, however, insufficient to demonstrate that hunting reputation is reflective of actual hunting ability. Hill & Kintigh (2009), although they did not directly assess measures of hunting reputation, have challenged the use of hunting reputation as a proxy measure for hunting ability due to the high variance in return rates for hunted animals. Hill & Kintigh (2009) have argued that, just as anthropologists may not be able to accurately estimate mean hunting success with <200-600 days of hunting return data, so too may a hunter's campmates find it hard to assess average success and, therefore, hunting ability. As Hadza most usually hunt alone, foods eaten out of camp (Berbesque et al., 2016) may further obfuscate hunting success; hunting reputation may reflect likelihood of bringing meat back to camp, rather than success and skill at killing wild animals. Furthermore, it is possible, as Kelly (2013; p.223) has highlighted, that individuals deemed better at hunting may simply be better known or better-thought-of, independent of their actual skill. Indeed, there was a weak but significant positive linear relationship between hunting reputation and 'renown', the number of people who knew each hunter. The direction of causality was here unclear, however, and it is not possible from the current results to say whether better hunters are better known by merit of their hunting ability or whether more well known individuals are more likely to be rated better hunters by dint of their popularity. The high level of agreement in our sample, therefore, does not, by itself, demonstrate that hunting reputation reflects true hunting ability.

Present results indicate that, at least when all hunters are included, assessments of the hunting ability of peers and campmates accurately predicted success on several independent or semi-independent foraging-related tasks. Hunting reputation is positively related to all of the variables expected to represent components of true hunting ability except for eyesight. Variation in these skills accounted for 44% of the variation in mean hunting scores. Draw strength, the ability to recognise animal sounds and archery skill each showed a significant positive relationship with hunting reputation. In concordance with the findings of Apicella (2014), draw strength was the best predictor of hunting reputation and appeared in every one of the potential best models derived from the data. Furthermore, when included in the same model, each of the three variables independently significantly predicted hunting reputation, and each appeared in the best model chosen using AIC (Section 3.3.1).

The lack of any significant relationship between hunting reputation and eyesight here is unexpected. Eyesight has been suggested, in previous discussions of hunting ability, to be an obvious component attribute (Blurton Jones, 2016, p.281). *A priori*, the reasoning is clear - in order to find and kill wild animals one must be able to see them. Current results indicate however that, although vision is clearly important in tracking prey, within the range of variation in the current dataset (no hunter was 'legally blind'), acute vision may be of subsidiary importance to other cues such as movement and calls.

As Hadza hunters usually hunt alone, others seldom have the opportunity to observe hunters using the particular skills measured in this study, especially knowledge of animal calls and aim. Furthermore, in interviews we asked not about specific hunting skills, but about hunting ability in general: 'Who here is the best hunter?' It is probable, therefore, that mean hunting scores are based on a general assessment of hunting ability, which is the product of several component hunting skills including, but not limited to, those skills measured in the study. The fact that hunting reputation is indeed significantly positively related to the three hunting skills measured in this study indicates that aggragated hunting reputation is a viable proxy for hunting ability in general.

The current data strongly indicate that, in aggregate, interviewees can accurately assess the true hunting ability of their current and former campmates but not the mechanism by which this is achieved. Given that daily return rates are so variable among the Hadza (Blurton Jones, 2016, p.30), the strength of the relationship between measured 'hunting skills' and hunting reputation may suggest that hunting returns are not the only cue used to assess hunting ability. It is unclear, however, what other cues interviewees may have used. Given the high level of agreement found here, it appears possible that opinions of the hunting ability of peers may be, in part, formed indirectly through discussions with others (for example through gossip). From the current data, we have no way of knowing whether assessments of hunting ability are the product of word-of-mouth 'indirect reputation' or are formed independently.

Hill & Kintigh (2009), although they did not test the claim directly, have also argued that, where a relationship between hunting reputation and hunting ability or return rate does occur, this may be due to a small number of good or bad hunters whose abilities are readily apparent. Present results only partially supported this view. When those nine hunters with hunting reputations >1.5 SD from the mean were excluded, age and aim at 10m no longer predicted hunting reputation, though animal recognition and strength remained significant predictors. When those 27 hunters with reputations above and below >1 SD from the mean a were removed (38.03% of hunters in this study) strength remained a significant predictor, while animal call recognition neared significance (p <0.1). These results suggest that interviewees were still able to distinguish between the relative skills of 'average' hunters, albeit less easily and with less agreement.

Overall, results show that, although not in complete agreement, independent interviewees provide a strong consensus about the hunting ability of their current and former campmates. Furthermore, Hill & Kintigh (2009) appear incorrect in their claim that people are unable to accurately distinguish between good and poor hunters and, taken in aggregate, mean hunting scores represented a viable proxy for measures of three different component hunting skills and probably also for general hunting ability across the study population.

4.2. Hunting Reputation and the Costly Signalling Hypothesis

The 'costly signalling hypothesis' predicts that men hunt food not primarily in order to provision themselves or their families, but to honestly 'signal particular hidden attributes' (Bliege Bird & Smith, 2005, p.221) and, in doing so, gain status and favourable attention from potential allies or mates (Hawkes, 1991; Hawkes et al., 1991, 1997). The extent to which hunting represents a 'cost' to the hunter and his family remains hotly debated (Wood & Marlowe, 2013; Hawkes et al., 2014; Wood & Marlowe, 2014), although to a certain extent this point is moot; cost is not a necessary condition of honest signalling (Maynard Smith & Harper, 2003, p.45). It is possible for honest signals to be cost free (e.g. Davies & Halliday, 1978; Taylor et al., 2000) and the emphasis placed by those in the human sciences on the demonstration of cost has been criticised (Grose, 2011; Számadó, 2011). The costly signalling hypothesis of human hunting is, however, supported by the fact that various measures of hunting success are, in a number of forager groups, well correlated with a variety of measures of reproductive success (Smith, 2004; Gurven et al., 2006).

Testing the costly signalling hypothesis further is made difficult by the fact that it is seldom specified what the hunting of widely redistributed food items serves to signal. Bliege Bird & Power (2015) have argued that hunting might serve as a signal of pro-sociality, which is unrelated to skill as at finding and killing wild animals given equal luck and effort (hunting ability). More papers, however, suggest that hunters signal qualities related to 'their hunting prowess' (p.59 Hawkes & Bird, 2002) and the component skills that contribute to it, for example 'ethological knowledge, visual acuity [or] stamina' (Smith & Bliege Bird, 2005, p.142). This is also often linked to the idea that women may 'attempt to pass on any genetic component of hunting skill to their own sons by choosing good hunters as mates' (Kaplan & Hill, 1985, p.133). The prediction we test here is put most clearly by Bliege Bird & Smith (2005), who suggest that 'hunting more successfully or productively' is an activity with an 'unambiguously ranked outcome that depend[s] upon participant skill or other hidden qualities' (p.233).

If hunting is, indeed, an honest signal of qualities related to hunting ability, it is necessary that hunting reputation is A) widely agreed upon and B) reflective of the component hunting skills that contribute to hunting ability, of which we measure a subset. If it is difficult, due, for example, to the obfuscating influence of daily variance ('luck') upon hunting acquisition rates, to accurately assess the hunting true ability of one's neighbours and peers (Hill & Kintigh, 2009), then hunting would be an ineffective signal of such qualities as it would not reliably convey any hidden information.

Given that hunting is generally solitary among the Hadza, campmates must assess hunting ability in aggregate and have little opportunity to assess component hunting skills independently. For this reason, although we here measure only a subset of the component hunting skills assumed to contribute to hunting ability, it appears probable that the patterns observed in these data are generalisable to those component hunting skills which we have not measured also.

As detailed above (Section 4.1), current results are at least ostensibly consistent with the costly signalling hypothesis. Results indicated generally high levels of agreement between independent interviewees. Furthermore, and contrary to the predictions of Hill & Kintigh (2009), the hunting reputation measure was correlated with several probable hunting skills, and interviewees could predict the hunting skills of 'middling' hunters, albeit with diminished accuracy. Although hunting ability is the result of many component skills, only a small subset of which were captured in the current study, current results indicate that hunting ability is, to an extent, known and agreed upon.

We can also comment on whether it is plausible that hunting might signal those specific qualities measured in the current study - strength, aim and knowledge of animal calls. Although strength may be readily accurately visually assessed (Fink et al., 2007; Sell et al., 2009), it is probable that knowledge of animal calls as well as aim are cryptic in everyday Hadza life. Aim requires developed fine motor skill, which may bring an adaptive advantage (Vashro & Cashdan, 2015). Furthermore, aim may be a desirable characteristic in potential allies and, given the lethality of the bow (Bingham, 2000; Churchill & Rhodes, 2009), is certainly a relevant characteristic in potential rivals. Knowledge of animal calls could be related to an individual's ability to learn, retain and assimilate information. Several authors have stressed the importance of learning in forager environments (Richerson & Boyd, 2000; Gurven et al., 2006; Wells & Stock, 2007; Kaplan et al., 2009; Boyd et al., 2011) and an ability to learn is, for this reason, probably an adaptively relevant trait of interest to prospective mates. Perhaps more directly, Hadza women themselves frequently cite 'intelligence' as among the most important traits of a potential husband, fourth after hunting ability itself, 'character' and physical appearance (Marlowe, 2004).

We show here that hunting could, therefore, potentially signal two qualities that are both otherwise cryptic and adaptively relevant/desirable in a potential partner. However, the fact that it *could* does not necessarily mean that it does. Honest signals in the zoological literature, although often costly, are usually efficient means of signalling information in a short time. Many honest signals (e.g. Andersson, 1992; Kodric-Brown, 1998; Johnsen et al., 2003; Hagen et al., 2004; Loyau et al., 2007) are visual and can be assessed by signal recipients immediately. Those signals that involve courtship rituals (Leal, 1999; Kotiaho, 2000; Langkilde et al., 2005; Murai & Backwell, 2006) are usually less than ten minutes in duration. It is rare to find a proposed instance of honest signalling which takes more than an hour to interpret by a signal recipient. The best-known exception to this trend, nest guarding and food-sharing in Arabian babblers, which occurs over the span of several months during a breeding season, Wright (1997, 2007) has argued, is most probably not a means of signalling at all and is better explained by inclusive fitness.

By contrast, hunting reputations may take years to build in human communities and are not immediately apparent to unfamiliar individuals. Although reputation was significantly positively correlated with hunting skills in the current study, the measured hunting skills accounted for only 44% of variation in the reputation measure. As we measured only a subset of the component skills of hunting, this is to be expected, and it is likely that measuring and incorporating additional hunting skills into the model would increase the fit, even when adjusting for the increased number of parameters. Nonetheless it is hard to imagine, given the high daily variance in food acquisitions noted by many (Cashdan, 1985; Winterhalder, 1986; Layton, 2005; Hill & Kintigh, 2009), that any hunting reputation measure would ever be wholly free from noise and error. There may be simpler ways to show off one's aim, depth of environmental knowledge or other capabilities which do not require the mean 6.3 hours that an adult Hadza man spends out of camp each day in search of food (Berbesque et al., 2016). And this logic should also apply to those component hunting skills (e.g. stealth, spotting or tracking) not measured in the current study as well. Though here shown to be effective, hunting, may, therefore, not be the most parsimonious method of signalling and the status of hunting as a 'handicap' or 'honest signal' *sensu* Zahavi (1975) and Maynard Smith & Harper (2003) is here called into question.

Furthermore, the fact that interviewees were less well able to predict the hunting skills of middling hunters raises some questions about the payoffs for individual hunters. Among the Hadza, although there is some variance in hunting effort, most men go hunting on most days. If hunting is primarily a means of signalling, and not primarily a subsistence strategy, it is unclear why any but the best hunters continue to hunt regularly. Despite a similar degree of hunting effort, middling hunters signal less effectively; their hunting skills were less well-predicted by their hunting reputation and more subject to noise and error. Furthermore, less well-reputed hunters signal their relative lack of ability more effectively and, were hunting solely a means of signalling, would do better by hunting less regularly in order to obfuscate their lack of skill. Although inter-individual variance in hunting effort appears low (most men spend most days hunting out of camp, and hunt for a broadly similar amount of time) it is possible that some men spend slightly more time foraging than others. It would be enlightening to assess whether there existed a relationship between hunting reputation and time spent hunting and this question may be fertile grounds for further research.

A final issue raised by this study is that, if hunting reputation is correlated with other qualities such as strength which are possible to assess independent of hunting, the noted associations between hunting reputation and reproductive success may be confounded (Smith, 2004; Apicella, 2014). Apicella (2014) has shown that hunting reputation predicts reproductive success even when controlling for strength. However, insufficient data on RS were collected in the current study to attempt to replicate this result. Furthermore, it is impossible to rule out the possibility that some unmeasured confounding variable, not captured in either the current study or Apicella's (2014) might be responsible for the observed relationship (Smith, 2004) between proximate measures of hunting ability and RS.

5. Conclusions

The use of hunting reputation as a proxy for hunting ability has elsewhere been criticised and Kelly (2013) and Hill & Kintigh (2009) have argued that hunting reputation measures might be unreliable.

In this study we introduced a novel measure of hunting reputation which involved asking interviewees to rank an entire bounded sample of hunters. This method improved upon the previously used 'nominations' method as it captured differences in hunting reputations between individuals at all levels. The method improved upon the 'within-camp ranking methods' as results were generalisable across camps and less subject to bias and sampling error. The method has two drawbacks. First it is time-consuming. Second, because rankings are relative to only those hunters within the sample, results cannot be generalised across the whole population. Nonetheless the advantages of the current method outweigh the disadvantages and those planning future studies specifically focussing on hunting reputation should consider following a similar protocol.

Furthermore, the results demonstrated that previous concerns about the efficacy of using reputation based measures as a proxy for true hunting ability are largely unfounded. Interviewees showed a high level of agreement which suggests A) that they have a similar understanding of the task they were being asked to complete and B) that they were assessing hunters using a broadly similar set of criteria. Furthermore, hunting reputation results were significantly positively related to at least three component hunting skills. Given that these skills are probably seldom directly observed by interviewees and are not tightly correlated, these results suggest that the aggregated hunting reputation measure was a reliable proxy for hunting skills in general.

Lastly, results were not, at first appearance, inconsistent with the costly signalling hypothesis and demonstrated that hunting effort could indeed signal several probably otherwise cryptic component hunting skills and related traits with some degree of accuracy. However, the measure was noisy enough to cast doubt on the efficiency of hunting reputation as a heuristic for those skills measured in the current study and, we infer, for other characteristics related to hunting ability also. Of course, this depends upon what quality hunting might actually signal. We have only captured a subset of potential candidates and if this idea is to be tested further, it important that proponents of the costly signalling hypothesis make more specific predictions about what exact skills or capabilities hunting serves to make clear. However, were hunting truly a means of 'showing off' otherwise cryptic capabilities related to hunting, such as the skills we have measured, we argue there might be other ways to do this which, while still being condition-honest, were subject to less error on the part of signal recipients and did not involve more than six hours per day spent foraging, in most cases, alone with no observers.

Acknowledgements and Funding Sources

This work was supported by Robinson College Cambridge, the Leakey Foundation, the Smuts Memorial Fund, the Ruggles-Gates Fund of the Royal Anthropological Institute, the Cambridge Department of Archaeology and Anthropology, the Cambridge Centre for African Studies, the Anthony Wilkin Fund and the Ridgeway-Venn Fund.

In addition, we would like to thank all those who have supported us at all stages of this project. We would especially like to highlight the contributions of Trevor Hawkes and Enrico Crema, for their help with our analysis.

Data Availability

As certain participants in the study are well known by their reputations as hunters, in the interests of protecting participants anonymity we have opted not to make these data freely available online. Those interested in accessing the data underlying this paper should contact the corresponding author using the e-mail address provided.

References

Aiello, L. C., & Wheeler, P. (1995). The Expensive-Tissue Hypothesis: The Brain and the Digestive System in Human and Primate Evolution. *Current Anthropology*, *36*, 199. doi:10.1086/204350.

Alvard, M. S., & Gillespie, A. (2004). Good Lamalera Whale Hunters Accrue Reproductive Benefits. In M. S. Alvard (Ed.), *Socioeconomic Aspects of Human Behavioral Ecology* (pp. 225–247). Bingley: Emerald Group Publishing Ltd volume 23 of *Research in Economic Anthropology*. (1st ed.). doi:10.1016/S0190-1281(04)23009-8.

Andersson, S. (1992). Female preference for long tails in lekking Jackson's widowbirds: experimental evidence. *Animal Behaviour*, *43*, 379–388. doi:10.1016/S0003-3472(05)80098-3.

Apicella, C. L. (2014). Upper-body strength predicts hunting reputation and reproductive success in Hadza hunter-gatherers. *Evolution and Human Behavior*, *35*, 508–518. doi:10.1016/j.evolhumbehav.2014.07.001.

Apicella, C. L., Crittenden, A. N., & Tobolsky, V. A. (2017). Evolution and Human Behavior Hunter-gatherer males are more risk-seeking than females , even in late childhood. *Evolution and Human Behavior*, *38*, 592–603. URL: http://dx.doi.org/10.1016/j.evolhumbehav.2017.01.003. doi:10.1016/j.evolhumbehav.2017.01.003.

Bagshawe, F. (1925). The Peoples of the Happy Valley (East Africa) The Aboriginal Races of Kondoa Irangi. Part II: The Kangeju. *Journal of the Royal African Society*, *24*, 117–139. doi:10.1093/afraf/101.404.427.

Berbesque, J. C., Wood, B. M., Crittenden, A. N., Mabulla, A., & Marlowe, F. W. (2016). Eat first, share later: Hadza Hunter-gatherer men consume more while foraging than in central places. *Evolution and Human Behavior*, *37*, 1–6. doi:10.1016/j.evolhumbehav.2016.01.003.

Bingham, P. M. (2000). Human Evolution and Human History: A Complete Theory. *Evolutionary Anthropology*, *9*, 248–257. doi:10.1002/1520-6505(2000)9:6<248::aid-evan1003>3.0.co;2-x.

Bird, R. (1999). Cooperation and conflict: The behavioral ecology of the sexual division of labor. *Evolutionary Anthropology*, *8*, 65–75. doi:10.1002/(SICI)1520-6505(1999)8:2<65::AID-EVAN5>3.0.CO;2-3.

Bliege Bird, R., & Power, E. A. (2015). Prosocial signaling and cooperation among Martu hunters. *Evolution and Human Behavior*, *36*, 389–397. doi:10.1016/j.evolhumbehav.2015.02.003.

Bliege Bird, R., & Smith, E. A. (2005). Signaling Theory, Strategic Interaction, and Symbolic Capital. *Current Anthropology*, *46*, 221–248. doi:10.1086/427115.

Blurton Jones, N. G. (2016). *Demography and Evolutionary Ecology of Hadza Hunter-Gatherers*. Cambridge: Cambridge University Press.

Blurton Jones, N. G., & Marlowe, F. W. (2002). Selection for delayed maturity. *Human Nature*, *13*, 199–238.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 10918–10925. doi:10.1073/pnas.1100290108.

Cashdan, E. A. (1985). Coping with Risk: Reciprocity Among the Basarwa of Northern Botswana. *Man*, *20*, 454–474. doi:10.2307/2802441.

Churchill, S., & Rhodes, J. (2009). The evolution of the human capacity for "killing at a distance": the human fossil evidence for the evolution of projectile weaponry. In J.-J. Hublin, & M. P. Richards (Eds.), *The Evolution of Hominin Diets* chapter 15. (pp. 201–210). Dordrecht: Springer Netherlands. (1st ed.).

Davies, N. B., & Halliday, T. R. (1978). Deep croaks and fighting assessment in toads Bufo bufo. *Nature*, *274*, 683–685. doi:10.1038/274683a0.

Domínguez-Rodrigo, M., Pickering, T. R., Semaw, S., & Rogers, M. J. (2005). Cutmarked bones from Pliocene archaeological sites at Gona, Afar, Ethiopia: Implications for the function of the world's oldest stone tools. *Journal of Human Evolution*, *48*, 109–121. doi:10.1016/j.jhevol.2004.09.004.

Durst, W., Peters, T., & Wilhelm, B. (2011). Testing acuity and contrast vision under standardised lighting conditions. *British Journal of Ophthalmology*, *95*, 1506–1508. doi:10.1136/bjophthalmol-2011-300353.

Fink, B., Neave, N., & Seydel, H. (2007). Male Facial Appearance Signals Physical Strength to Women. *American Journal of Human Biology*, *19*, 82–87. doi:10.1002/ajhb.

Grose, J. (2011). Modelling and the fall and rise of the handicap principle. *Biology and Philosophy*, *26*, 677–696. doi:10.1007/s10539-011-9275-1.

Gurven, M., Kaplan, H., & Gutierrez, M. (2006). How long does it take to become a proficient hunter? Implications for the evolution of extended development and long life span. *Journal of Human Evolution*, *51*, 454–470. doi:10.1016/j.jhevol.2006.05.003.

Gurven, M., & von Rueden, C. (2006). Hunting, social status and biological fitness. *Social Biology*, *53*, 81–99. doi:10.1080/19485565.2006.9989118.

Hagen, S. B., Debeausse, S., Yoccoz, N. G., & Folstad, I. (2004). Autumn coloration as a signal of tree condition. *Proceedings of the Royal Society B: Biological Sciences*, *271 Suppl*, S184–S185. doi:10.1098/rsbl.2003.0126.

Hawkes, K. (1991). Showing off: Tests of an Hypothesis About Men's Foraging Goals. *Ethology and Sociobiology*, *12*, 29–54. doi:10.1016/0162-3095(91)90011-E.

Hawkes, K., & Bird, R. B. (2002). Showing Off, Handicap Signaling, and the Evolution of Men's Work. *Evolutionary Anthropology*, *11*, 58–67.

Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (1991). Hunting income patterns among the Hadza: big game, common goods, foraging goals and the evolution of the human diet. *Philosophical Transactions of the Royal Society B: Biological sciences*, *334*, 243–250; discussion 250–251. doi:10.1098/rstb.1991.0113.

Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (2001). Hunting and Nuclear Families: Some Lessons from the Hadza about Men's Work. *Current Anthropology*, *42*, 681–709. doi:10.1086/322559.

Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (2014). More Lessons from the Hadza about Men's Work. *Human Nature*, *25*, 596–619. doi:10.1007/s12110-014-9212-5.

Hawkes, K., O'Connell, J. F., & Rogers, L. (1997). The behavioral ecology of modern hunter-gatherers, and human evolution. *Trends in Ecology and Evolution*, *12*, 29–32. doi:10.1016/S0169-5347(96)10060-4.

Hill, K., & Hurtado, A. M. (1996). *Ache life history: The ecology and Demography of a Foraging People*. Hawthorne: Aldine de Gruyter.

Hill, K., Kaplan, H., & Hawkes, K. (1993). On Why Male Foragers Hunt and Share Food. *Current Anthropology*, *34*, 701. doi:10.1086/204213.

Hill, K., & Kintigh, K. (2009). Can Anthropologists Distinguish Good and Poor Hunters? Implications for Hunting Hypotheses, Sharing Conventions, and Cultural Transmission. *Current Anthropology*, *50*, 369–378. doi:10.1086/597981.

Howell, N. (1979). *Demography of the Dobe !Kung*. (2nd ed.). New York: Hawthorne.

Johnsen, A., Delhey, K., Andersson, S., & Kempenaers, B. (2003). Plumage colour in nestling blue tits: Sexual dichromatism, condition dependence and genetic effects. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 1263–1270. doi:10.1098/rspb.2003.2375.

Kaplan, H., Gurven, M., & Winking, J. (2007). An Evolutionary Theory of Human Life Span: Embodied Capital and the Human Adaptive Complex. In V. L. Bengtson, D. Gans, N. M. Putney, & M. Silverstein (Eds.), *Handbook of Theories of Aging* chapter 3. (pp. 39–61). New York: Springer. (2nd ed.).

Kaplan, H., & Hill, K. (1985). Hunting ability and reproductive success among male Ache foragers: Preliminary results. *Current Anthropology*, *26*, 131–133.

Kaplan, H. S., Hooper, P. L., & Gurven, M. (2009). The evolutionary and ecological roots of human social organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3289–3299. doi:10.1098/rstb.2009.0115.

Keeffe, J. E., Lovie-Kitchin, J. E., Maclean, H., & Taylor, H. R. (1996). A simplified screening test for identifying people with low vision in developing countries. *Bulletin of the World Health Organization*, *74*, 525.

Kelly, R. L. (2013). *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*. (3rd ed.). Cambridge: Cambridge University Press.

Kodric-Brown, A. (1998). Sexual dichromatism and temporary color changes in the reproduction of fishes. *American Zoologist*, *38*, 70–81. doi:10.1093/icb/38.1.70.

Koster, J. (2010). Informant Rankings via Consensus Analysis. *Current Anthropology*, *51*, 257–258. doi:10.1086/651073.

Kotiaho, J. S. (2000). Testing the assumptions of conditional handicap theory: costs and condition dependence of a sexually selected trait. *Behavioral Ecology and Sociobiology*, *48*, 188–194. doi:10.1007/s002650000221.

Langkilde, T., Schwarzkopf, L., & Alford, R. A. (2005). The Function of Tail Displays in Male Rainbow Skinks (Carlia jarnoldae). *Journal of Herpetology*, *39*, 325–328.

Layton, R. (2005). Are Immedaite Return Strategies Adaptive? In T. Widlok, & W. Tadesse (Eds.), *Property and Equality Volume 1: Ritualisation, Sharing, Egalitarianism* (pp. 130–150). New York: Berghahn.

Leal, M. (1999). Honest signalling during prey-predator interactions in the lizard Anolis cristatellus. *Animal Behaviour*, *58*, 521–526. doi:10.1006/anbe.1999.1181.

Loyau, A., Gomez, D., Moureau, B., Thery, M., Hart, N. S., Jalme, M. S., Bennett, A. T., & Sorci, G. (2007). Iridescent structurally based coloration of eyespots correlates with mating success in the peacock. *Behavioral Ecology*, *18*, 1123–1131. doi:10.1093/beheco/arm088.

Marlowe, F. W. (1999). Showoffs or Providers? The Parenting Effort of Hadza Men. *Evolution and Human Behavior*, *20*, 391–404. doi:10.1016/S1090-5138(99)00021-5.

Marlowe, F. W. (2000). The patriarch hypothesis. *Human Nature*, *11*, 27–42.

Marlowe, F. W. (2004). Mate preferences among Hadza hunter-gatherers. *Human Nature*, *15*, 365–376.

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology*, *14*, 54–67. doi:10.1002/evan.20046.

Marlowe, F. W. (2007). Hunting and Gathering: The Human Sexual Division of Foraging Labor. *Cross-Cultural Research*, *41*, 170–195. doi:10.1177/1069397106297529.

Marlowe, F. W. (2010). *The Hadza: Hunter-Gatherers of Tanzania*. Los Angeles: University of California Press.

Marlowe, F. W., Berbesque, J. C., Wood, B., Crittenden, A., Porter, C., & Mabulla, A. (2014). Honey, Hadza, hunter-gatherers, and human evolution. *Journal of Human Evolution*, *71*, 119–128. doi:10.1016/j.jhevol.2014.03.006.

Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. (1st ed.). Oxford: Oxford University Press. doi:10.1080/00219266.1997.9655528.

McPherron, S. P., Alemseged, Z., Marean, C. W., Wynn, J. G., Reed, D., Geraads, D., Bobe, R., & Béarat, H. A. (2010). Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature*, *466*, 857–60. doi:10.1038/nature09248.

Murai, M., & Backwell, P. R. Y. (2006). A conspicuous courtship signal in the fiddler crab Uca perplexa: Female choice based on display structure. *Behavioral Ecology and Sociobiology*, *60*, 736–741. doi:10.1007/s00265-006-0217-x.

Murdock, G. P. (1967). Ethnographic Atlas: A Summary. *Ethnology*, *6*, 109–236.

Obst, E. (1912). Von Mkalama ins land der Wakindiga. *Mitteilungen der Geographischen Gesellschaft in Hamburg*, *26*, 3–45.

Rabinovich, R., Gaudzinski-Windheuser, S., & Goren-Inbar, N. (2008). Systematic butchering of fallow deer (Dama) at the early middle Pleistocene Acheulian site of Gesher Benot Ya'aqov (Israel). *Journal of Human Evolution*, *54*, 134–149. doi:10.1016/j.jhevol.2007.07.007.

Richerson, P. J., & Boyd, R. (2000). Climate, culture and the evolution of cognition. In C. Heyes, & L. Huber (Eds.), *The Evolution of Cognition* chapter 18. (pp. 329–345). Cambridge: MIT Press. (1st ed.).

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 575–584.

Smith, E. A. (2004). Why do good hunters have higher reproductive success? *Human Nature*, *15*, 343–364. doi:10.1007/s12110-004-1013-9.

Smith, E. A., & Bliege Bird, R. (2005). Costly Signaling and Cooperative Behavior. In H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* chapter 4. (pp. 115–148). Cambridge: MIT Press. doi:10.1525/aa.2007.109.2.380.

Smith, E. A., Bliege Bird, R., & Bird, D. W. (2003). The benefits of costly signaling: Meriam turtle hunters. *Behavioral Ecology*, *14*, 116–126. doi:10.1093/beheco/14.1.116.

Számadó, S. (2011). The cost of honesty and the fallacy of the handicap principle. *Animal Behaviour*, *81*, 3–10. doi:10.1016/j.anbehav.2010.08.022.

Taylor, P. W., Hasson, O., & Clark, D. L. (2000). Body postures and patterns as amplifiers of physical condition. *Proceedings of the Royal Society B: Biological Sciences*, *267*, 917–922.

Thieme, H. (1997). Lower Palaeolithic hunting spears from Germany. *Nature*, *385*, 807–810. doi:10.1038/385807a0.

Vashro, L., & Cashdan, E. (2015). Spatial cognition, mobility, and reproductive success in northwestern Namibia. *Evolution and Human Behavior*, *36*, 123–129. doi:10.1016/j.evolhumbehav.2014.09.009.

Wells, J. C. K., & Stock, J. T. (2007). The Biology of the Colonizing Ape. *Yearbook of Physical Anthropology*, *50*, 191–222. doi:10.1002/ajpa.20735.

Winterhalder, B. (1986). Diet choice, risk, and food sharing in a stochastic environment. *Journal of Anthropological Archaeology*, *5*, 369–392. doi:10.1016/0278-4165(86)90017-6.

Wood, B. M., & Marlowe, F. W. (2013). Household and kin provisioning by hadza men. *Human Nature*, *24*, 280–317. doi:10.1007/s12110-013-9173-0.

Wood, B. M., & Marlowe, F. W. (2014). Toward a Reality-Based Understanding of Hadza Men's Work. *Human Nature*, *25*, 620–630. doi:10.1007/s12110-014-9218-z.

Wright, J. (1997). Helping-at-the-nest in Arabian babblers: Signalling social status or sensible investment in chicks? *Animal Behaviour*, *54*, 1439–1448. doi:10.1006/anbe.1997.0563.

Wright, J. (2007). Cooperation theory meets cooperative breeding: exposing some ugly truths about social prestige, reciprocity and group augmentation. *Behavioural Processes*, *76*, 142–148. doi:10.1016/j.beproc.2007.01.017.

Zahavi, A. (1975). Mate selection - A selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214. doi:10.1016/0022-5193(75)90111-3.

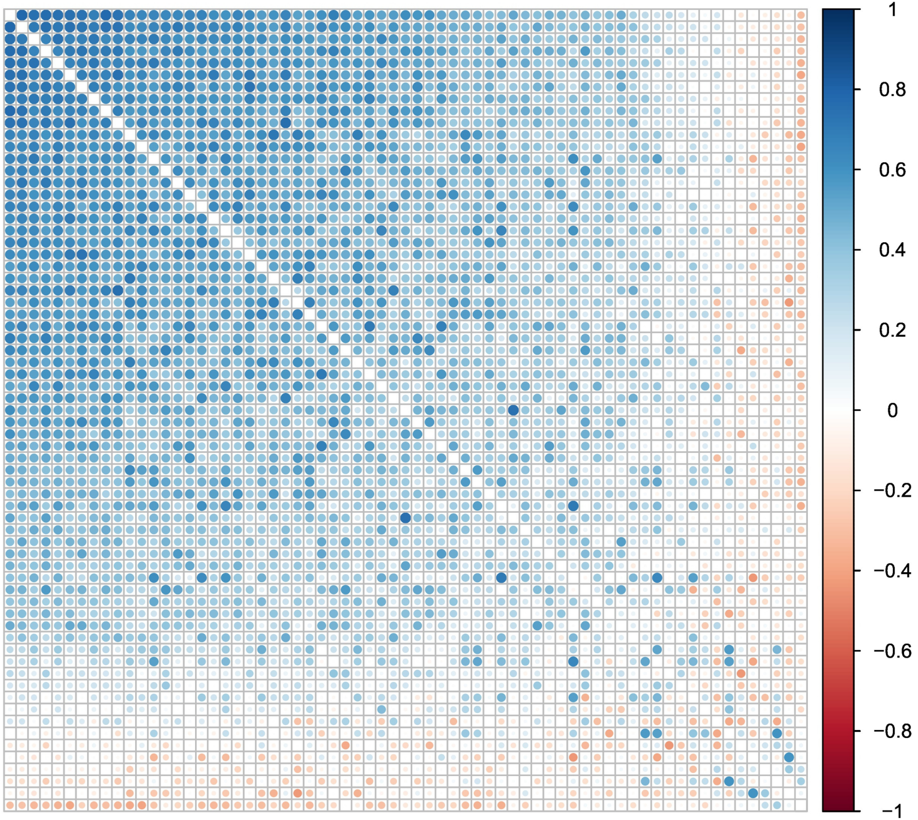


Figure 1

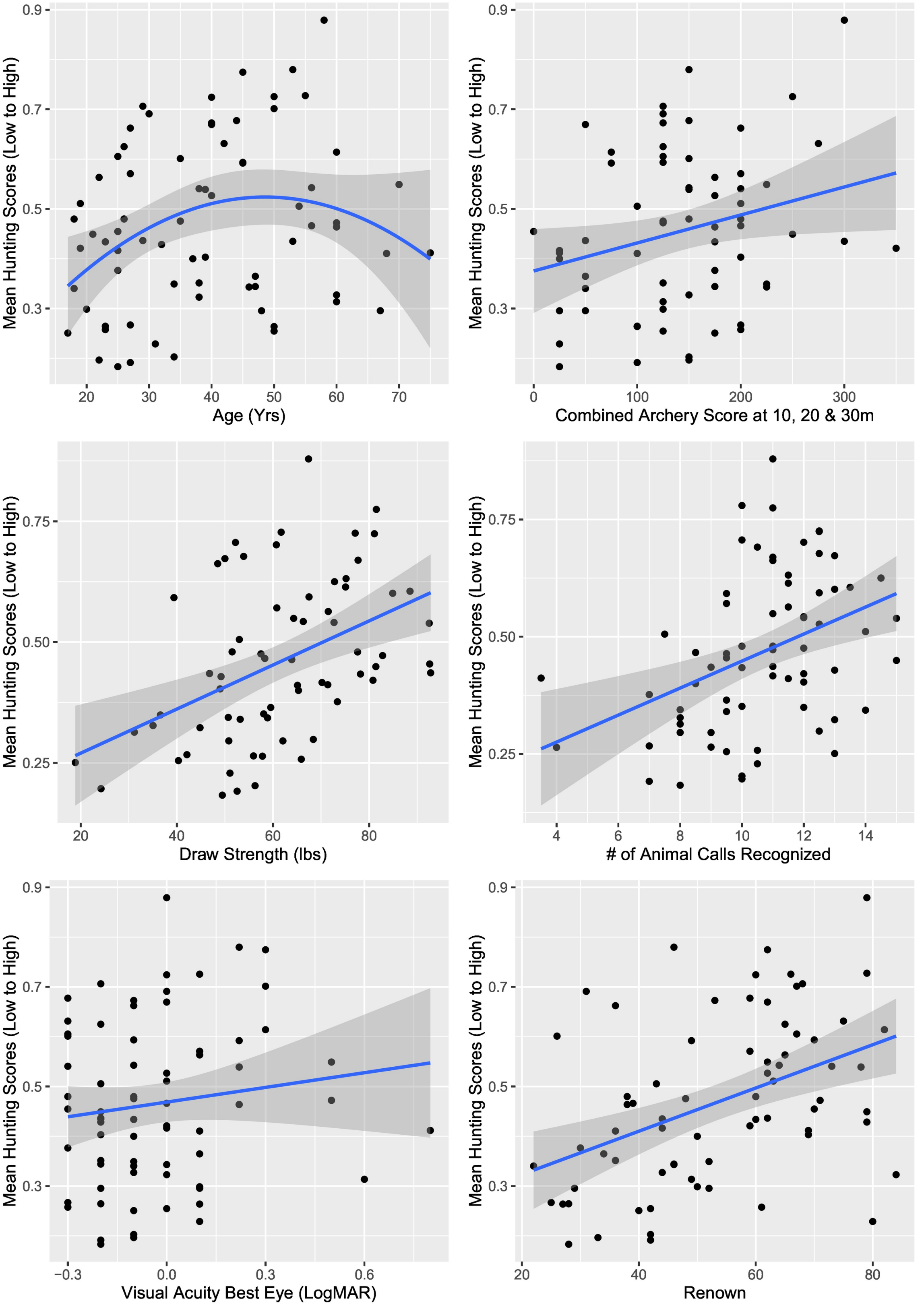

Mean Hunting Scores (Low to High)
Age (Yrs)
Combined Archery Score at 10, 20 & 30m
Draw Strength (lbs)
# of Animal Calls Recognized
Visual Acuity Best Eye (LogMAR)
Renown

Figure 2